

The very wide-field gzK galaxy survey – I. Details of the clustering properties of star-forming galaxies at $z \sim 2$

Shogo Ishikawa^{1,2*}, Nobunari Kashikawa^{1,2}, Jun Toshikawa², and Masafusa Onoue^{1,2}

¹*Department of Astronomy, School of Science, SOKENDAI (The Graduate University for Advanced Studies), Mitaka, Tokyo 181-8588, Japan*

²*Optical and Infrared Astronomy Division, National Astronomical Observatory of Japan, Mitaka, Tokyo 181-8588, Japan*

8 February 2016

ABSTRACT

We present the results of clustering analysis on the $z \sim 2$ star-forming galaxies. By combining our data with data from publicly available archives, we collect g -, z_B/z_- , and K -band imaging data over 5.2 deg^2 , which represents the largest area BzK/ gzK survey. We apply colour corrections to translate our filter-set to those used in the original BzK selection for the gzK selection. Because of the wide survey area, we obtain a sample of 41,112 star-forming gzK galaxies at $z \sim 2$ (sgzKs) down to $K_{AB} < 23.0$, and determine high-quality two-point angular correlation functions (ACFs). Our ACFs show an apparent excess from power-law behaviour at small angular scale ($\theta \lesssim 0.01^\circ$), which corresponds the virial radius of a dark halo at $z \sim 2$ with a mass of $\sim 10^{13} M_\odot$. We find that the correlation lengths are consistent with the previous estimates over all magnitude range; however, our results are evaluated with a smaller margin of error than that in previous studies. The large amount of data enables us to determine ACFs differentially depending on the luminosity of the subset of the data. The mean halo mass of faint sgzKs ($22.0 < K \leq 23.0$) was found to be $\langle M_h \rangle = (1.32_{-0.12}^{+0.09}) \times 10^{12} h^{-1} M_\odot$, whereas bright sgzKs ($18.0 \leq K \leq 21.0$) were found to reside in dark haloes with a mass of $\langle M_h \rangle = (3.26_{-1.02}^{+1.23}) \times 10^{13} h^{-1} M_\odot$.

Key words: cosmology: observations — dark matter — galaxies: evolution — galaxies: formation — large-scale structure of universe — surveys

1 INTRODUCTION

Exploring galaxy formation and evolution is one of the most important issues in modern astronomy. The structure formation model suggesting that more massive objects in the Universe were rapidly assembled, and star formation activities were completed in an early epoch is known as “downsizing” (e.g., Cowie et al. 1999); however, it is not clear when and how the specific precesses in this evolutionary model occurred. The physical processes between baryons in the dark matter halo are so complex that it is not known how the luminous objects that we are able to observe directly were formed in the highly dense dark matter halo.

To approach these issues, it is important to determine properties of galaxies, including the dark halo mass, the stellar mass, the star-formation rate (SFR), and the morphology over cosmic time, and to compare these quantities with theoretical models. The dark halo mass of galaxies is a particularly important parameter to trace the mass assembly history of galaxies because, according to the Λ CDM

model, dark haloes grow monotonically with cosmic time by merging, irrespective of the baryon processes. However, measuring the mass of the dark halo is not straightforward. One of the most effective methods to determine the mass of dark haloes is to use the galaxy clustering strength. Various studies have revealed the dark halo mass in the distant Universe (e.g., Kashikawa et al. 2006; Hildebrandt et al. 2009; Jose et al. 2013) as well as in the local Universe (e.g., Zehavi et al. 2005; Zheng et al. 2007); however, this method requires a large number of samples in order to measure the clustering strength of galaxies, which has been a problem for $z \sim 2$ galaxies.

The distinctive spectral features of galaxies at $z \sim 2$ are red-shifted from optical wavelengths to near-infrared (NIR) wavelengths. Wide-field NIR observation using ground-based telescopes is complicated because of the limited physical size of the NIR detectors and (with the exception of some atmospheric windows) poor sky transparency, as well as strong OH airglow emission lines and thermal emission. For this reason, wide-field galaxy survey data are lacking at $z \sim 2$ (this commonly referred to as the “redshift desert”). Recent observations have shown that $z \sim 2$ is an important era in galaxy

* E-mail: shogo.ishikawa@nao.ac.jp

formation and evolution. For example, star formation in galaxies peaked at $1 < z < 2$ (e.g., Hopkins & Beacom 2006; Burgarella et al. 2013), galaxy mass rapidly assembled at $1 < z < 3$ (e.g., Arnouts et al. 2007; Pérez-González et al. 2008; Santini et al. 2009; Ilbert et al. 2010), and the number density of QSOs peaked at $z \sim 2$ (e.g., Palanque-Delabrouille et al. 2013).

A powerful method to select unbiased data with high completeness $z \sim 2$ galaxies has been proposed, namely the “BzK selection” technique (Daddi et al. 2004). The BzK selection method requires only B -, z -, and K -band photometric data, and enables us to select both star-forming BzKs (hereafter sBzKs) and passively evolving BzKs (hereafter pBzKs) simultaneously in a single colour-colour diagram. However, a precise clustering analysis of BzK galaxies requires a large number of galaxy samples and has not yet been performed due to the difficulties in obtaining wide-field K -band imaging data.

A number of studies have attempted to reveal the clustering properties of BzK galaxies. Kong et al. (2006) constructed BzK galaxy samples over 1200 arcmin^2 . They conducted the clustering analysis of sBzKs and pBzKs and concluded that pBzKs are more strongly clustered than sBzKs, suggesting that the morphology–density relation seen in the local Universe must already have been in place at $z \sim 2$. Hayashi et al. (2007) carried out a deep BzK galaxy survey in the Subaru Deep Field to investigate the properties of faint sBzKs and reported that the clustering strength of sBzKs depends on the K -band luminosity. McCracken et al. (2010) succeeded in constructing a large number of BzK galaxy samples in the COSMOS field over 1.9 deg^2 down to $K < 23.0$ and carried out a precise clustering analysis. Wide survey area and multiwavelength data from the COSMOS survey enabled them to determine the comoving correlation length for both sBzKs and pBzKs, respectively. More recently, Bielby et al. (2014) reported the clustering properties of $z \sim 2$ galaxies by splitting the galaxy samples into star-forming galaxies and passive galaxies based on the photometric redshift and galaxy colour. They revealed that passive galaxies are more strongly clustered than are star-forming galaxies; however, very massive star-forming galaxies ($M_* \sim 10^{11} M_\odot$) exhibited comparable clustering strength to passive galaxies. Their clustering analysis of star-forming galaxies at $z \sim 2$ was, however, confined to massive galaxies because of a shortage of galaxy samples at $z \sim 2$. Béthermin et al. (2014) succeeded in calculating the dark halo mass of sBzKs for wide mass range ($10^{10.4} M_\odot < M_* < 10^{11.4} M_\odot$) using both galaxy clustering and HOD analyses in the COSMOS field. The dark halo masses determined using both methods were in good agreement and were consistent with the predictions of semi-analytic models. They calculated the dark halo mass at $z \sim 2$; however, the clustering analysis used had a poor signal-to-noise (S/N) ratio, especially at the large angular scale ($\theta \gtrsim 0.1^\circ$), which made it difficult to determine the dark halo mass accurately.

In this paper, we report a largest-ever wide-field galaxy survey over $\sim 5 \text{ deg}^2$ down to $K_{AB} < 23.0$ using g -band data instead of B -band data (we term our sample galaxies “gzKs”) and performed the clustering analysis of star-forming gzKs (hereafter sgzKs) with a significantly higher S/N ratio than has been reported previously. Because of

the large number of sgzK samples, we are able to divide our sgzK samples into subsamples according to the K -band luminosity, and we investigate the clustering properties of each subsample. Our very wide survey area makes it possible to constrain the dark halo mass because we were able to determine an accurate angular correlation function at large angular scales, which is essential to determining the dark halo mass accurately. In this paper (paper I), we describe the measurement of the angular correlation function of star-forming galaxies at $z \sim 2$ based on this large sample, and the dark halo mass is determined using angular correlation functions. Additionally, we carried out a halo occupation distribution (HOD) analysis on the sgzKs, and we determine the distribution of galaxies in the dark haloes for each K -band limiting magnitude, which will be described in paper II.

The framework of this paper is organized as follows. In Section 2, we give details of the data that were used as well as the method employed to construct the K -selected and sgzK catalogues. We describe the method used to correct for differences in the filter transmissions between our gz_BK filters and the VLT- BzK filters (Daddi et al. 2004). In Section 3, we describe the method to conduct clustering analysis using the angular correlation function. Taking advantage of the large number of sgzK samples, we resample the sgzK samples into subsamples according to the K -band magnitudes to investigate the clustering properties. The results of our clustering analysis of sgzKs are discussed in Section 4, and we give a conclusion in Section 5. All magnitudes and colours are in the AB system. Throughout this paper, we assume flat lambda cosmology ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$), the Hubble constant is $h = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1} = 0.7$, and the normalization of the matter power spectrum is $\sigma_8 = 0.8$. Assuming these cosmological parameters, the age of the Universe at $z = 2$ is $\sim 3.22 \text{ Gyr}$, and 1 arcsec corresponds to $8.37h^{-1} \text{ kpc}$ in the comoving scale.

2 DATA

2.1 Photometric Data

We collected wide-field imaging data by using both our own data and publically available archival data. Photometric data describing the K -band were obtained from the archives of the United Kingdom Infra-Red Telescope (UKIRT) Deep Sky Survey (UKIDSS; Lawrence et al. 2007) in the Deep Extragalactic Survey (DXS). The UKIDSS data were acquired using the Wide Field Camera (WFCAM), which is composed of four $2,048 \times 2,048$ pixel detectors, with a separation of 12.8 between detectors. Each detector covers $13.65' \times 13.65'$ of the sky at a pixel scale of 0.187 pixel^{-1} . The central wavelength and the full-width at half-maximum (FWHM) of the K -band of the WFCAM are $\lambda_c = 2.2 \mu\text{m}$ and $\text{FWHM} = 0.34 \mu\text{m}$, respectively.

We obtained z_B -band photometric data during the course of another observational program (Kashikawa et al. 2015) using the Subaru Telescope/Suprime-Cam from June 22 to June 24, 2009. Suprime-Cam is a wide-field imaging instrument covering $34' \times 27'$ of the sky at a pixel scale of 0.22 pixel^{-1} . The z_B -filter ($\lambda_c = 8,842 \text{ Å}$, $\text{FWHM} = 689 \text{ Å}$) is a custom-made filter that divides the SDSS z -band filter at $9,500 \text{ Å}$ and their bluer one (Shimasaku et al.

Table 1. Limiting magnitudes and covering areas of each band

	limiting magnitude (3σ , $2''\phi$ in AB)	area (deg ²)
g	26.08	5.2
z_B	25.55	3.2
z	25.36	2.0
K	23.31	5.2

2005). One of our observation field “VIMOS4” is centered on ($22^h20^m00^s, +00^\circ42'00''$; J2000.0) and covers ~ 3 deg² (Lawrence et al. 2007). Here, 3 deg² of the VIMOS/DXS field was covered by 13 Suprime-Cam field-of-views (FoVs). We adopted a common dithering circle pattern comprising a full cycle of dithering consisting of 5 pointings. The total integration time was 1,800 seconds per FoV. The sky condition was very good, with a seeing size of $0.''6$. Photometric calibration was carried out using the spectroscopic standard stars GD153 and Feige110. The data were reduced using the pipeline software package SDFRED (Yagi et al. 2002; Ouchi et al. 2004a). The package includes bias subtraction, flat fielding, a correction for image distortion due to the prime focus, PSF matching, sky subtraction, and mosaicking.

We also retrieved z -band photometric data from the SMOKA data archive server (Baba et al. 2002) to extend the survey field to 5.2 deg². These data satisfy the conditions that they connect continuously to our z_B -band field, and the limiting magnitude is comparable to our z_B -band data. The z -band data were obtained using the Subaru Telescope/Suprime-Cam.

We retrieved g -band images, which cover the entire field where K and z_B/z images are available, using the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS; Gwyn 2011) archival data instead of using B -band. The CFHTLS data were acquired using the CFHT MegaCam, which covers $57.'6 \times 56.'4$ of the sky at a pixel scale of $0.''187$ pixel⁻¹, where the central wavelength and FWHM of the g -band of MegaCam are $\lambda_c = 4,870\text{\AA}$ and $\text{FWHM} = 1,450\text{\AA}$, respectively.

We collected (g, z_B, K) bands of imaging data over 3.2 deg² and (g, z, K) bands over 2.0 deg². The survey field of this study covers 5.2 deg², comprising the above two regions with different imaging datasets.

Fang et al. (2012) showed that star-forming and passive galaxies at $z \sim 2$ are also able to be selected using the ($g - z$) and the ($z - K$) colours, which is similar to BzK colour selection. They constructed sgzK and pgzK samples by modifying the original BzK colour-selection criteria and validated their criteria using the stellar population synthesis model (Bruzual & Charlot 2003). We applied a similar gzK selection process to our data, although colour correction from our ($g - z_B$) data to the VLT- $(B - z)$ data is required, as only the z_B -band is available.

2.2 K -selected Catalogue

The g - and z_B/z -band images were matched to the geometry of the K -band image using the IRAF task `geomatch`. Object

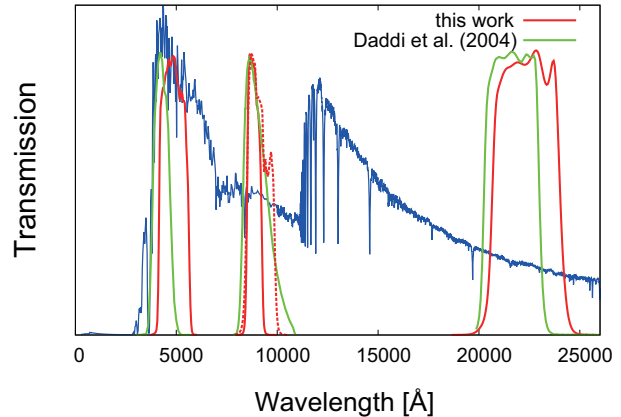


Figure 1. Differences of the filter transmission profiles between our $gzBK$ -filters (red) and the VLT- BzK filters (green) used in Daddi et al. (2004). Our filters are composed of CFHT/MegaCam g -band, Subaru/Suprime-Cam z_B/z -band, and UKIRT/WFCAM K -band. Suprime-Cam z -band is presented by the dashed line. The blue line represents the typical SED of sgzK galaxies.

detection and photometry were carried out using SExtractor, version 2.8.6 (Bertin & Arnouts 1996). Following object detection in the K -band, we used SExtractor in “double-image mode” to measure the object fluxes in the g - and z_B/z -bands with $2''$ aperture photometry. The size of the unit image was $\sim 4,170 \times 4,170$ pixels, which corresponds to $\sim 27'.8 \times 27'.8$. We carefully conducted masking on the low-S/N regions, such as near the edges of the images and close to the saturated objects, to remove objects with an uncertain flux. The number of masks was 690 over the field, and the effective survey area was 4.8 deg². We detected 339,581 objects in our field after masking.

The limiting magnitudes were measured with a 3σ , $2''$ aperture in AB magnitude. We used the IRAF task `limitmag` in SDFRED to determine the limiting magnitudes and applied averaged data as limiting magnitudes in each band. Table 1 lists a summary of the limiting magnitudes and the areas covered by each band. These limiting magnitudes were almost identical (δ mag ~ 0.08) over the entire field.

2.3 Band Corrections

We used the BzK selection method to select $z \sim 2$ galaxy samples; however, the filter set of our data differed slightly from the VLT- BzK filter set (Daddi et al. 2004). Figure 1 shows a comparison of the transmission of our $gzBK$ -filters with that of VLT- BzK filters. When adapting our $gzBK$ data to the VLT- BzK data, careful correction for differences in the filters is essential to adequately select $z \sim 2$ galaxies and to compare the results.

Daddi et al. (2004) used the K_s -band (Skrutskie et al. 2006), whereas we used K -band imaging data from the UKIDSS archive. We used the following conversion equation (Carpenter 2001) between these filters:

$$K_s = K + 0.002 + 0.026 \times (J - K). \quad (1)$$

J -band imaging data were also retrieved from the UKIDSS archive, as both K -band photometric data and J -band magnitudes were available. To simplify the notation, we hereafter

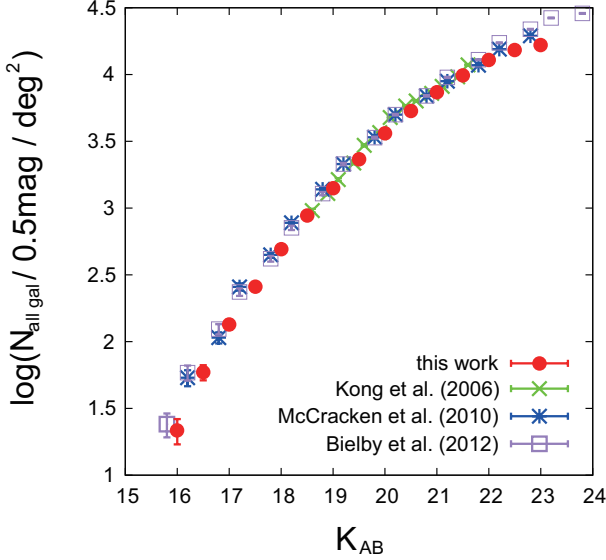


Figure 2. Number count of the K -band selected galaxies after the correction of the detection completeness in our VIMOS4 field (red circles). Green, blue, and purple symbols represent the results of Kong et al. (2006), McCracken et al. (2010), and Bielby et al. (2012), respectively.

describe the K_s -band magnitudes converted from K -band magnitudes using equation (1) as “ K ”. Following conversion of the K -band magnitudes, we limited our K -selected catalogue to $K = 23.0$, which corresponds to a 70% completeness cut (see Section 3.3). We extracted all galaxies from our catalogue using the BzK diagram (see Section 2.4), and sorted them in magnitude bins from $K = 16.0$ to $K = 23.0$ with $\delta K = \pm 0.25$.

Figure 2 shows a comparison of the number counts of our K -selected galaxies with previously reported studies. Our K -selected catalogue is consistent with previous works.

The BzK selection technique selects $z \sim 2$ galaxies in the colour-colour diagram of $(B - z)$ versus $(z - K)$. We attempted to carry out colour conversion from $(g - z_B)$, $(g - z)$, and $(z_B - K)$ to $(B - z)$ and $(z - K)$ following K -band magnitude conversion. First, to correct for differences between the z_B -band and the z -band, we simultaneously measured the z_B - and the z -band fluxes for the stars in the ~ 0.22 deg² region, where photometric data are available for both, and derived the following conversion equation:

$$(z - K) = 0.940 \times (z_B - K) + 0.00330 \quad (2)$$

using a least-squares regression method.

In the same way, we determined the following expressions for conversion:

$$(g - z) = 0.900 \times (B - z)_{\text{Subaru}} - 0.120 \quad (3)$$

and

$$(g - z_B) = 0.961 \times (g - z) + 0.115, \quad (4)$$

respectively. We employed the public photometric catalogue of the COSMOS survey, in which g -, B -, and z -band photometries are available, to derive the equation (3).

McCracken et al. (2010) introduced empirical relations to translate $(B - z)$ colour derived by the Subaru/Suprime-

Cam data to $(B - z)$ colour derived by the VLT for objects with $(B - z)_{\text{Subaru}} < 2.5$,

$$(B - z)_{\text{VLT}} = 1.0833 \times (B - z)_{\text{Subaru}} + 0.053 \quad (5)$$

and with $(B - z)_{\text{Subaru}} \geq 2.5$,

$$(B - z)_{\text{VLT}} = (B - z)_{\text{Subaru}} + 0.27. \quad (6)$$

Using these empirical relations and equation (3), we obtain the relations that convert from the Subaru- $(g - z)$ colour to the VLT- $(B - z)$ colour for objects with $(g - z) < 2.13$,

$$(B - z)_{\text{VLT}} = 1.20 \times (g - z) + 0.198 \quad (7)$$

and with $(g - z) \geq 2.13$,

$$(B - z)_{\text{VLT}} = 1.11 \times (g - z) + 0.404. \quad (8)$$

To convert from the Subaru- $(g - z_B)$ colour for objects with $(g - z_B) < 2.16$,

$$(B - z)_{\text{VLT}} = 1.25 \times (g - z_B) + 0.0535 \quad (9)$$

and with $(g - z_B) \geq 2.16$,

$$(B - z)_{\text{VLT}} = 1.16 \times (g - z_B) + 0.271. \quad (10)$$

We validated these colour convention equations by using simulated galaxy spectral energy distribution (SED) models. The various SED models of galaxies were generated by varying physical properties such as the star-formation history, the SFR, the age, and the dust extinction using the stellar population synthesis models (Bruzual & Charlot 2003). We used the dust extinction law of Calzetti et al. (2000). We determined the $(g - z_B)$ and $(z_B - K)$ colours from those SEDs, and translated them to the $(B - z)$ and $(z - K)$ colours using the relations presented above, and then traced the redshift evolution on the colour-colour diagram. We found that star-forming galaxies and passively evolving galaxies at $1.4 \lesssim z \lesssim 2.5$ with our band corrections met the original selection criteria proposed by Daddi et al. (2004).

2.4 gzK Selection Method

We constructed the $z \sim 2$ star-forming galaxy samples by applying the BzK selection method. We term our BzK selection method the “gzK selection” method, and we select star-forming galaxy samples as sgzKs because we use g -band photometric data instead of B -band data.

The criterion used to select the sBzK galaxies proposed by Daddi et al. (2004) is given by

$$(z - K) \geq (B - z) - 0.2, \quad (11)$$

and that for pBzK is given by

$$(z - K) < (B - z) - 0.2 \cap (z - K) > 2.5. \quad (12)$$

We applied these original criteria after converting (g, z_B, z, K_s) photometric data to (B, z, K) photometric data using equations (2) and (7) ~ (10).

Figure 3 shows a colour-colour diagram for all the objects down to $K = 23.0$ in our VIMOS4 field, where the blue and red dots represent the sgzKs and pgzKs, respectively. The number of selected sgzKs was $N(\text{sgzKs}) = 41,112$, and the number of pgzKs was $N(\text{pgzKs}) = 1,313$. Our g -band photometry was too shallow to detect all of the pgzK galaxies, which had very red $(g - z)$ colour (at least $g - z > 2.7$);

therefore, we consider only the *sgzKs* galaxies. It should be noted that the *g*-band limiting magnitudes of non-detected objects in the *g*-band were replaced by the 2σ limiting magnitude in the *g*-band, whereas non-detected objects in the z_B/z -band were removed from our catalogue because their locations in the colour–colour diagram were uncertain.

Figure 4 shows a comparison of the number count between the *sgzK* and *sBzK* galaxies from our data and from previous studies. After correcting the completeness (see Section 3.2), our data were found to be consistent with almost all previous studies. Table 2 lists a summary of the number count of our *sgzKs*. Our results were in agreement with almost all previous studies and were in particularly good agreement with the data reported by McCracken et al. (2010). Our study and McCracken et al. (2010) are particularly in good agreement each other, while other previous studies have slightly higher number counts than our study. This might be caused by their small survey area.

On the other hand, the bright end of our *sgzK* number count showed an excess compared with these data reported by Blanc et al. (2008) and Hartley et al. (2008). Our very wide survey field enabled us to detect quite rare bright *sgzKs* at $K \lesssim 19$.

In our colour–colour diagram, there are some anomalous features. First, in the *sgzK* region, the sharp vertical line at $(B - z, z - K) \sim (1.0, 2.6)$ to $\sim (1.0, 3.3)$, which is originated from the non-detected objects in the *g*-band, can be seen. The z/z_B -band magnitudes of the objects on this line are so close to the limiting magnitude that the $(B - z)$ colour is nearly constant. We note that the $(B - z)$ colour was converted from original $(g - z)$ and $(g - z_B)$ colours by equation (7) and (9). In addition, a coherent line is seen from $(B - z, z - K) \sim (1.0, 2.6)$ to $\sim (4.0, 0.0)$ in the diagram. This line is caused by the limiting magnitude in *K*-band, making the constant $(B - K)$ colour.

As shown in Figure 4, the number count of our *sgzK* sample is consistent with the previous results, suggesting that these anomalous features do not largely affect the *sgzK* selection; however, *g*-band faint *pgzKs*, which show stronger correlation, especially at the small angular scale, than *sgzKs* (e.g., McCracken et al. 2010; Zehavi et al. 2011), could come into the *sgzK* sample. To estimate the contamination of *g*-band faint *pgzKs* in our *sgzK* sample, we use the *K*-selected photometric redshift (photo-*z*) catalogue in the COSMOS field (Muzzin et al. 2013), which is approximately 1.0 magnitude deeper than our sample. We extracted *sgzKs* and *pgzKs* from COSMOS photo-*z* catalogue by applying our filter correction to that catalogue and the same colour selection criteria. Then we evaluated how many *pgzKs* are incorrectly selected as *sgzKs* by replacing *g*-band magnitudes by the 2σ limiting magnitude when the objects were fainter than the 2σ limiting magnitude. The contamination fractions of each limiting magnitude were found to be $f_{22 < K < 23}^{\text{pgzK}} = 8.8\%$, $f_{21 < K < 22}^{\text{pgzK}} = 8.8\%$, and $f_{18 < K < 21}^{\text{pgzK}} = 6.0\%$, respectively. Therefore, we concluded that the contamination from *pgzKs* to our *sgzK* sample is not substantial.

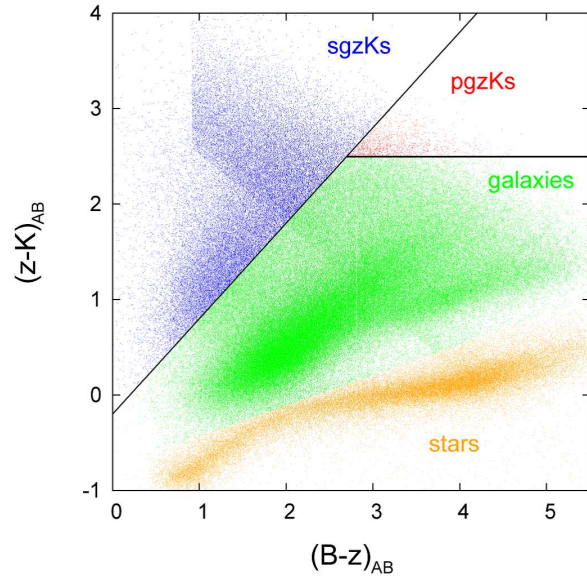


Figure 3. A colour–colour diagram of our VIMOS4 field. The blue, red, green and orange dots represent *sgzK* galaxies, *pgzK* galaxies, other galaxies and stars, respectively.

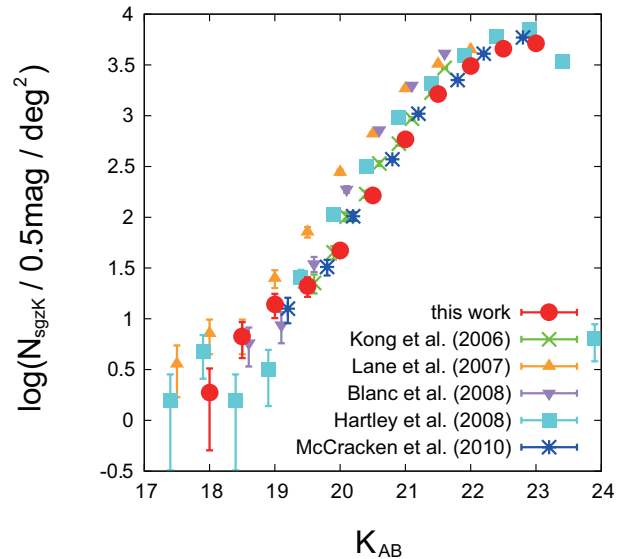


Figure 4. Comparison of the number counts between our *sgzK* galaxies (red) after completeness correction and the *sBzK* galaxies of the previous studies. Green crosses, orange uptriangles, purple downtriangles, cyan squares and blue stars correspond to Kong et al. (2006), Lane et al. (2007), Blanc et al. (2008), Hartley et al. (2008) and McCracken et al. (2010), respectively.

3 CLUSTERING ANALYSIS OF SGZK GALAXIES

3.1 Angular Correlation Function

The angular correlation function (hereafter ACF) is a useful indicator of the clustering properties of galaxies. The ACF $\omega(\theta)$ is defined as the probability P that a pair of galaxies exist within solid angles $\delta\Omega_1, \delta\Omega_2$ with a separation angle θ on the plane of the sky, and is given by

Table 2. The sgzK number count of each magnitude bin after completeness correction.

K_{AB}	N_{sgzK}	$\log(N_{\text{sgzK}}) \text{ deg}^{-2}$
18.0	9	0.273
18.5	32	0.825
19.0	67	1.143
19.5	101	1.323
20.0	225	1.673
20.5	790	2.217
21.0	2,801	2.766
21.5	7,844	3.213
22.0	14,789	3.489
22.5	21,875	3.659
23.0	24,704	3.712

$$\delta P = \bar{n}[1 + \omega(\theta)]\delta\Omega_1\delta\Omega_2, \quad (13)$$

where \bar{n} represents the average surface density of galaxies.

We used the standard estimator proposed by Landy & Szalay (1993), i.e.,

$$\omega(\theta) = \frac{DD - 2DR + RR}{RR}, \quad (14)$$

where DD , DR , and RR are the number of galaxy–galaxy, galaxy–random, and random–random pairs within the angular range $\theta - \delta\theta/2 < \theta < \theta + \delta\theta/2$, respectively, and are normalized to the number of all galaxy–galaxy, galaxy–random, and random–random pairs. We note that θ is in units of degrees.

We set the angular bin from $\log(\theta) = -3.6$ to $\log(\theta) = 0.2$, with a bin size $\delta\log(\theta) = 0.2$ to determine the ACF for sgzKs galaxies. We generated 300,000 random points homogeneously over the survey field in order to reduce the Poisson error. The random points were distributed to avoid the masked region, which we used when obtaining sgzK galaxy samples.

The ACF is generally given in power-law form, i.e.,

$$\omega(\theta) = A_\omega\theta^{1-\gamma}. \quad (15)$$

We applied a fixed power-law gradient at large scales with $\gamma = 1.8$ to compare the results with those of previous studies (e.g., Hayashi et al. 2007; Blanc et al. 2008).

It is well known that the ACF reported by Landy & Szalay (1993) is biased due to the “integral constraint” caused by the limitations of the observation field. The integral constraint is given by

$$C = \frac{1}{\Omega^2} \int_{\Omega} d\Omega_1 d\Omega_2 \omega(\theta). \quad (16)$$

We can estimate the integral constraint as follows:

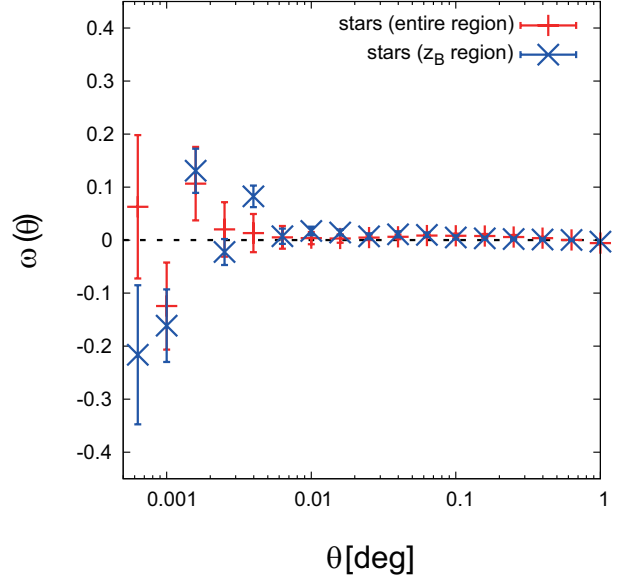
$$C = \frac{\sum_i RR_i \omega(\theta_i)}{\sum_i RR_i}, \quad (17)$$

provided that the ACF has already been obtained for each angular bin (Roche & Eales 1999). The bias-corrected ACF is given by subtracting the integral constraint, i.e.,

$$\omega(\theta) = A_\omega(\theta^{1-\gamma} - C). \quad (18)$$

The integral constraint in the VIMOS4 field was $C = 1.18$.

The error in the ACF was estimated using the bootstrap method as follows. We randomly resampled our sgzK

**Figure 5.** The ACFs of stars (red lines, entire survey region; blue crosses, z_B -band covered region) in our survey field satisfying $18.0 \leq K \leq 23.0$.

galaxies, allowing for redundancy and calculated the ACF at each step, repeating this 300 times. The uncertainty in each angular bin was determined from the root mean square of all of the bootstrap steps.

To confirm the validity of the homogeneity, we check the ACF of stars in our sample. The distribution of stars can be regarded as random distribution; therefore, their ACF is expected to be almost zero at all angular scale. Figure 5 shows the ACF of stars. The amplitude of each angular bin is approximately zero, indicating that our sample are almost selected homogeneously.

3.2 Subsamples

We divided our sample into subsamples according to the K -band luminosity to investigate the luminosity dependence of the clustering properties. Two subsample sets were created: a cumulative luminosity resampling subset and a differential luminosity resampling subset. The former was for use in comparing the results with previous studies, in which cumulative luminosity subsamples were created due to the small limited of sBzK data. We divided our large sgzK sample into five subsamples, i.e., $18.0 \leq K \leq 21.0$, $18.0 \leq K \leq 21.5$, $18.0 \leq K \leq 22.0$, $18.0 \leq K \leq 22.5$, and $18.0 \leq K \leq 23.0$. The sample number of each subsample are summarized in Table 3.

With the cumulative subsample set, however, it was difficult to describe the clustering properties of sgzKs by their K -band luminosity because cumulative resampling dilutes the differences among luminosity subsamples. Taking advantage of the large number of samples, we also divided our sample differentially according to the luminosity to assess the K -band luminosity dependence of sgzK clustering. We divided our sgzK samples into three subsamples based upon their K -band luminosity, i.e., $18.0 \leq K \leq 21.0$, $21.0 < K \leq 22.0$, and $22.0 < K \leq 23.0$, which allowed us to investigate how

clustering and the physical properties of *sgzK* galaxies depend on the *K*-band luminosity.

3.3 Redshift Distributions and Completeness

We estimated the redshift distribution of our *gzK*-selected sample by applying the *gzK* selection method (including the filter corrections) to the *K*-selected COSMOS photo-*z* catalogue (Muzzin et al. 2013). We assumed that the redshift distributions estimated from the COSMOS field ($\sim 1.9 \text{ deg}^2$) were the same as our large ($\sim 5.2 \text{ deg}^2$) field and that the cosmic variance was negligible.

We used **z_m2**, which usually provides the most feasible z_p , as described in the EAZY manual (Brammer et al. 2008) in the COSMOS photo-*z* catalogue. We extracted galaxies with a *K*-band magnitude in the range $18.0 \leq K \leq 23.0$ at $0 < z_p < 3$ and constructed a galaxy catalogue containing 79,284 galaxies. It should be noted that we excluded the galaxies with a large z_p error, i.e., $(|z_p - z_m2| > 0.5)$, where z_p is the best-fit redshift determined using SED fitting. The number of excluded galaxies was 318.

We calculated the $(g - z_B)$ and $(z_B - K)$ colours using the best-fit SEDs and by estimating z_p for each galaxy, and we derived the *g*- and z_B -band magnitudes according to the *K*-band magnitude from the COSMOS photo-*z* catalogue. Artificial galaxies with assigned (g, z_B, K) band photometries were then randomly distributed as point sources on the *g*-, z_B -, and *K*-band images. A total of 6,000 distributed galaxies were randomly selected from the galaxy catalogue allowing for redundancy in each limiting magnitude. Source detection and *sgzK* colour selection were carried out as with our data, and we obtained the redshift distributions for each luminosity subsample by repeating these processes 100 times and averaging the results of all steps.

Figure 6 and 7 show the redshift distributions that satisfy the limiting magnitudes of our subsamples. The redshift range that satisfies the *sgzK* criterion is $1.4 \lesssim z \lesssim 2.5$, which indicates the validity of our colour corrections. Our redshift distributions are almost consistent with those of previous studies (e.g., McCracken et al. 2010; Fang et al. 2012), though these are slightly shifted to lower z . This could be caused by the differences in the filters and the slight shallowness in the *K*-band of our sample. It should be noted that the brightest subsample ($18.0 \leq K \leq 21.0$) may be contaminated from low- z galaxies. We corrected for the effects of this contamination in the ACF for this brightest subsample by multiplying $1/(1 - f_c)^2$, where f_c is the contamination fraction ($f_c \sim 0.2$), assuming that the contaminating sources are not clustered each other. It should be noted, however, that this assumption is strictly incorrect for contaminated low- z galaxies, whose clustering amplitudes are difficult to be estimated. Other subsamples exhibited less contamination with low/high- z galaxies ($f_c < 0.1$), and we did not apply a correction factor to these data.

The redshift distribution that we estimated using the above procedure shows the completeness of our sample. Figure 8 shows our sample completeness; we obtained completeness of 70% with a limiting magnitude of $K_{AB} = 23.0$.

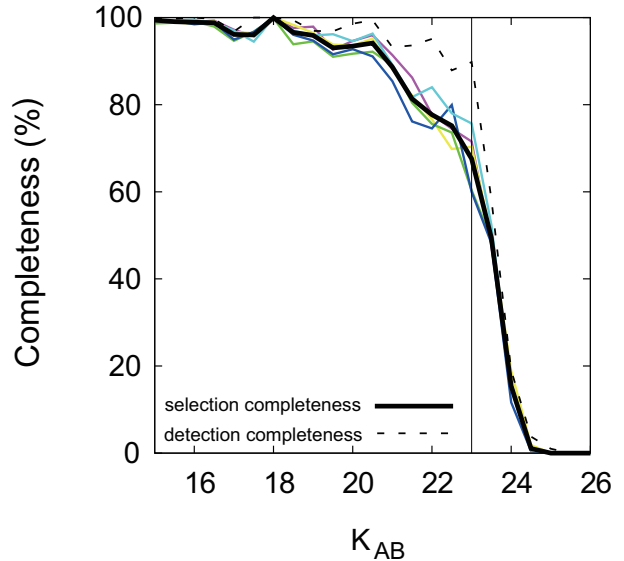


Figure 8. Completeness of our sample. The coloured thin lines represent the results in different regions and the black solid thick line is the average of the results in 5 regions. The dotted line is the detection completeness, which represents how inputted *sgzK*s were detected as objects. We applied $18.0 \leq K \leq 23.0$ magnitude cut to our sample, corresponding to approximately 70% completeness cut.

4 RESULTS AND DISCUSSIONS

We compare the results of our clustering analysis based on the cumulative luminosity subsampling with those of previous studies. We then show the results of the differential luminosity subsamples to reveal the luminosity dependence of the clustering properties.

4.1 Cumulative Luminosity Subsample

4.1.1 Angular Correlation Functions of Cumulatively Resampled *sgzK* Galaxies

The upper panel of Figure 9 shows the ACFs of each limiting-magnitude subsample. Bright *sgzK* galaxies were more strongly clustered than were faint *sgzK*s, which is consistent with previous studies (e.g., Kong et al. 2006; Hayashi et al. 2007; McCracken et al. 2010). The bin sizes of bright *sgzK* subsamples ($18.0 \leq K \leq 21.0$, $18.0 \leq K \leq 21.5$) were doubled ($\delta \log(\theta) = 0.4$) to gain the S/N ratio. Because of the accurate ACFs due to the large number of *sgzK* samples, our ACFs show an apparent excess from the power law at small angular scale (i.e., $\theta \lesssim 0.01^\circ$). These characteristics are attributed to the so-called “1-halo term”, which comes from galaxy clustering in the same dark halo. $\theta \sim 0.01^\circ$ corresponds to $0.21 h^{-1} \text{ Mpc}$ at $z \sim 2$ at a physical scale and is comparable to the virial radius ($\sim r_{200}$) of a dark halo at $z = 2$ with a mass of $\sim 10^{13} M_\odot$, where the virial radius is $r_{200} \sim 0.12 h^{-1} \text{ Mpc}$. The virial radius of the dark halo was estimated using

$$r_{200} = \left(\frac{GM_{200}}{100H^2(z)} \right)^{1/3}, \quad (19)$$

where G is the gravitational constant, M_{200} is the virial mass of the dark halo (Carroll et al. 1992; Ferguson et al. 2004),

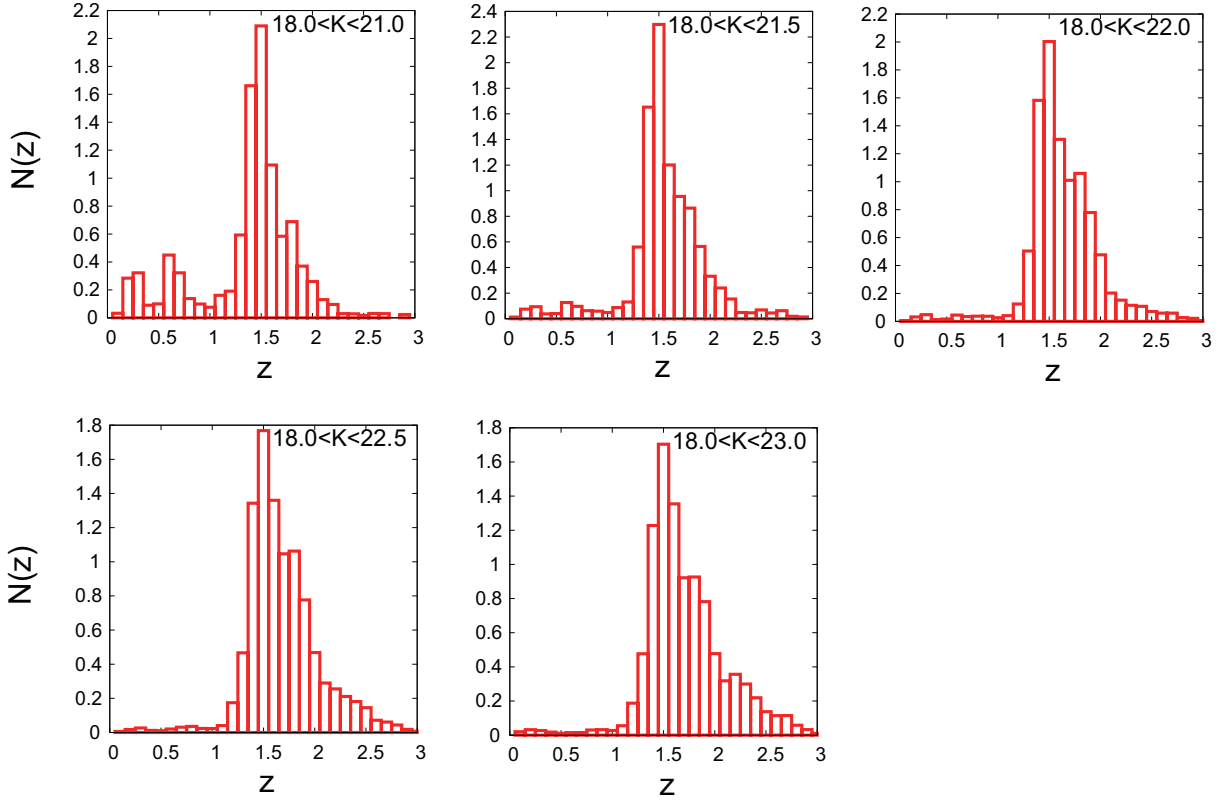


Figure 6. Redshift distributions of each cumulative subsample. The mean number of redshift bin with $\delta z = 0.1$ width is normalized to one. We extracted the galaxies at $0 < z < 3$ from the COSMOS K -selected catalogue (Muzzin et al. 2013) and applied gzK selection as we did on our own galaxy sample.

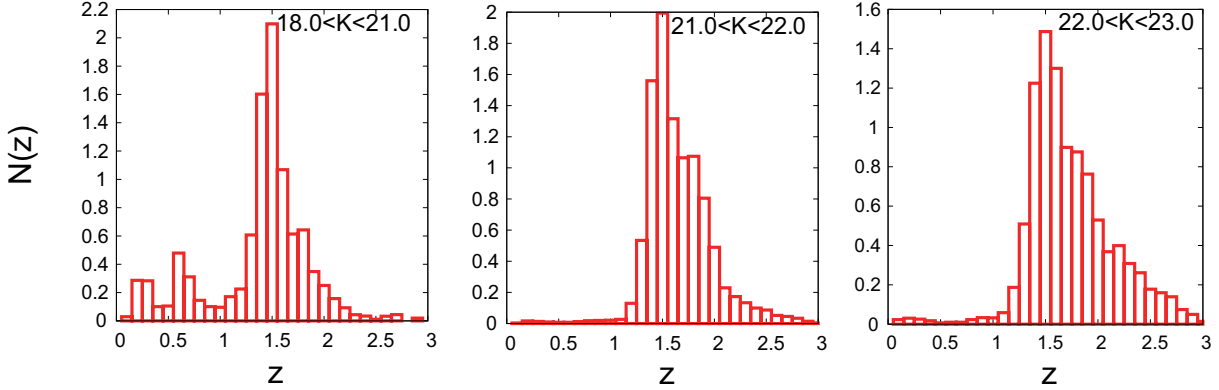


Figure 7. Redshift distributions of each differential subsample. The method to derive these redshift distributions is the same as Figure 6.

and $H(z)$ is the Hubble parameter with a redshift z as given by

$$H(z) = H_0 [\Omega_m(1+z)^3 + (1-\Omega_m-\Omega_\Lambda)(1+z)^2 + \Omega_\Lambda]^{1/2}. \quad (20)$$

On the other hand, the ACFs at large angular scale (i.e., $\theta \gtrsim 0.01^\circ$) were well approximated by a power law with an index of $\gamma = 1.8$, referred to as the “2-halo term”, which originates from galaxy clustering in different dark haloes. These two components can be accurately described by the halo occupation distribution (HOD) model, which predicts the galaxy distribution in the dark halo as a function of the

halo mass. We will describe the results of HOD analysis of our sgzK galaxies in our paper II.

We also show the model prediction of the dark matter ACF at $z \sim 2$ using the selection function of sgzKs with a limiting magnitude of $18.0 \leq K \leq 23.0$. This dark matter ACF was computed using the nonlinear Λ CDM power spectrum (Hamana et al. 2004, and the references therein).

The lower panel of Figure 9 shows the bias parameters of the sgzKs, which are defined as

$$b(\theta) = \sqrt{\frac{\omega_{\text{sgzK}}(\theta)}{\omega_{\text{DM}}(\theta)}}, \quad (21)$$

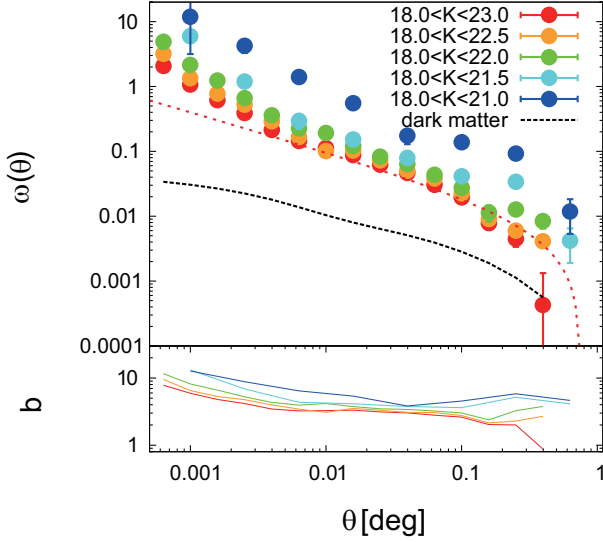


Figure 9. Upper panel: ACFs of the cumulatively resampled *sgzK* galaxies. The limiting magnitudes of each *sgzK* subsample are $18.0 \leq K \leq 23.0$, $18.0 \leq K \leq 22.5$, $18.0 \leq K \leq 22.0$, $18.0 \leq K \leq 21.5$, and $18.0 \leq K \leq 21.0$ (red, orange, green, cyan, and blue circles). We confirmed the fact that more bright *sgzK* galaxies show the strongly clustering, reported by Hayashi et al. (2007). The red-dotted line represents the result of single power-law fit to the ACF of the total sample at large angular scales to show the excess from a power law at small angular scales. The dotted line represents the model prediction of the dark matter ACF computed using the nonlinear power spectrum. Lower panel: The bias parameters of *sgzK*s. Bias parameter is defined as $b(\theta) = \sqrt{\omega_{\text{sgzK}}(\theta)/\omega_{\text{DM}}(\theta)}$.

where $\omega_{\text{sgzK}}(\theta)$ and $\omega_{\text{DM}}(\theta)$ are the amplitudes of the ACFs of *sgzK*s and dark matter at the angular scale θ , respectively. The values of our bias parameter were in the range $3 < b < 5$ at the large scale, which is consistent with the results of Blanc et al. (2008). The bias parameters were larger at small angular scales than at large angular scales, indicating the excess from the power law.

By applying the least χ^2 fitting of the single power law (equation 15) with fixed $\gamma = 1.8$, we obtained best-fit values of the amplitude of the ACF at 1° . We fitted the power law to the data for large angular scale ($\theta \gtrsim 0.01^\circ$) where the 1-halo term is negligible. Table 3 lists the resulting amplitudes of the ACF; we can see that the clustering amplitude is dependent on the *K*-band luminosity.

Our measurement of the amplitudes at $18.0 \leq K \leq 23.0$ was $A_\omega(1^\circ) = (3.33 \pm 0.09) \times 10^{-3}$, which is larger than the value of $(1.79 \pm 0.17) \times 10^{-3}$ reported by Hartley et al. (2008), as well as $(1.27 \pm 0.23) \times 10^{-3}$ reported by McCracken et al. (2010). For $18.0 \leq K \leq 22.0$, we found $A_\omega(1^\circ) = (3.83 \pm 0.14) \times 10^{-3}$, which is larger than the value of $(3.14 \pm 1.12) \times 10^{-3}$ reported by Blanc et al. (2008) and $(2.12 \pm 0.65) \times 10^{-3}$ by McCracken et al. (2010). This deviation may be due to the fact that the survey area of the previous studies was too small to provide a high-quality signal at the large scale ($\theta > 0.1^\circ$). The ACFs reported in most previous studies were truncated at $\theta < 0.1^\circ$ or declined due to the effects of integral constraints. With our results, however, which are based on wide-field data, we calculated the ACF over a wide angular scale of $0.01^\circ < \theta < 0.5^\circ$, which enabled

us to more accurately determine the amplitude, especially for the 2-halo term of the ACF. The amplitude of the ACF can be calculated from the large-scale galaxy clustering; the angular range was approximately $0.01^\circ \lesssim \theta \lesssim 0.1^\circ$, which makes it difficult to calculate the amplitude accurately if intermediate to large-scale clustering is not well determined.

In addition, Sato et al. (2014) pointed out that the clustering amplitude reported by McCracken et al. (2010) was weaker than those of the other studies. Sato et al. (2014) also presented the correlation functions, which is inconsistent with the result of McCracken et al. (2010), at the COSMOS field, though the origin of this discrepancy was unclear.

The results of our brightest three bins were in good agreement with those of Kong et al. (2006) although the error bars for their data were relatively large. Blanc et al. (2008) attributed the large amplitude reported by Kong et al. (2006) to the effects of cosmic variance; however, our results, where the survey field was more than 10 times larger than that of Kong et al. (2006), are less affected by cosmic variance. For this reason, we can calculate the amplitude of the ACF of *K*-bright *sgzK*s more accurately than Kong et al. (2006) were able to.

4.1.2 Clustering in Real-Space

The correlation amplitude of the ACF, A_ω , can be transformed into the three-dimensional correlation length by assuming a redshift distribution, where the correlation length corresponds to the three-dimensional clustering strength. The redshift distributions we used are described in Section 3.2 (see Figure 6 and 7).

The spatial two-point correlation function, $\xi(r)$, is described in power-law form as follows:

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \quad (22)$$

where normalization factor r_0 is termed the correlation length, and γ is the gradient of the power law. We assumed $\gamma = 1.8$, in a similar way as with the ACF.

The conversion between the ACF and the two-point correlation function was implemented using Limber's equation (Limber 1953). Provided that the redshift distribution is known, the amplitude of the ACF can be translated to the correlation length as follows (Peebles 1980; Efstathiou et al. 1991):

$$A_\omega(1\text{rad}) = r_0^\gamma \sqrt{\pi} \frac{\Gamma(\gamma - 1/2)}{\Gamma(\gamma/2)} \times \frac{\int_0^\infty F(z) D_A^{1-\gamma}(z) N(z)^2 g(z) dz}{[\int_0^\infty N(z) dz]^2}. \quad (23)$$

The term $A_\omega(1\text{rad})$ corresponds to the amplitude of the ACF at $\theta = 1$ radian, where Γ is the Gamma function, D_A is the angular diameter distance, and $N(z)$ is the redshift distribution, respectively. The function $g(z)$ depends on the cosmological parameters and is given as follows:

$$g(z) = \frac{H_0}{c} (1+z)^2 \sqrt{1 + \Omega_m z + \Omega_\Lambda [(1+z)^{-2} - 1]}. \quad (24)$$

$F(z)$ is a function describing the redshift evolution of the two-point correlation function $\xi(r)$, i.e.,

$$F(z) = (1+z)/(1+z_c)^{-(3+\epsilon)}, \quad (25)$$

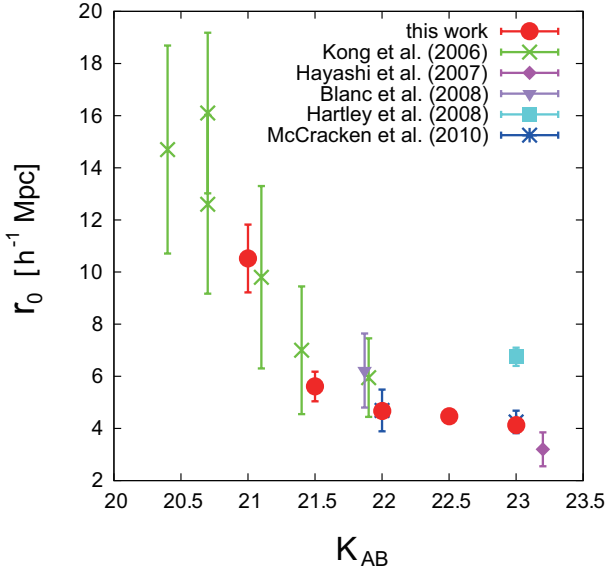


Figure 10. Comparison of our correlation lengths of sgzKs with previous studies. Our results (red filled circles) are consistent with previous studies over all magnitude range and measured with small error bars. All correlation lengths are in units of h^{-1} Mpc, where $h = 0.7$.

where z_c is the center of the redshift distribution. We assumed ϵ as $\epsilon = -1.2$ (see also Ouchi et al. 2004b).

The correlation lengths are listed in Table 3, and Figure 10 shows a comparison of our results with those of previous studies. We found that brighter sgzKs have larger correlation lengths, which indicates that brighter galaxies reside in more massive haloes and exhibit stronger clustering. It should be noted that because of the large sample size, our results are characterized by smaller error bars than those of previous studies. Our results also show excellent agreement with previous studies over all magnitudes, with the exception of Hartley et al. (2008). This discrepancy between the result of Hartley et al. (2008) and rest of the studies we compare may be caused by the inaccuracy of their sample selection (see also McCracken et al. 2010; Sato et al. 2014). As discussed in section 4.1.1, our ACFs have higher amplitudes than the previous studies, whereas our correlation lengths exhibit good agreement with the previous results. This may be due to the slight difference of the redshift distribution, that is, our sample has a redshift distribution shifted to lower- z than the previous studies. We confirmed that the correlation length becomes smaller when the redshift distribution shifts low- z .

The correlation length for sgzKs with $18.0 \leq K \leq 21.0$ were determined using the contamination-corrected ACF. Our correlation length of the brightest subsample is relatively large compared with the result of Kong et al. (2006), which was not corrected for contamination; however, as shown in Figure 6, the brightest subsample was more contaminated by the low/high- z galaxies and should be corrected accordingly. The difference in the correlation lengths can be explained by the lack of correction for contamination in the data reported by Kong et al. (2006).

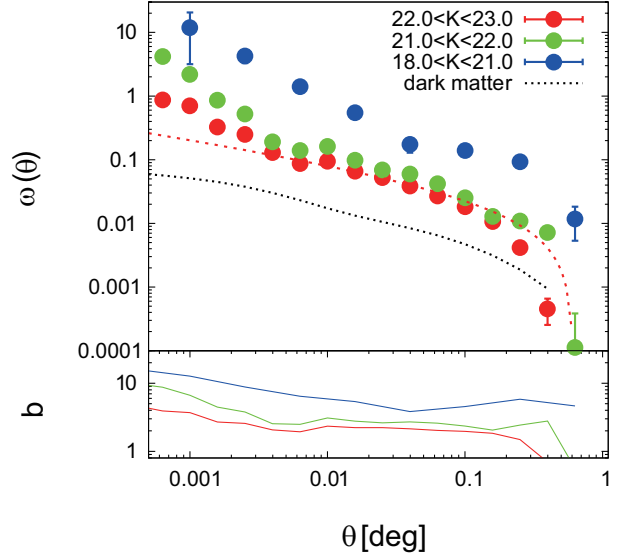


Figure 11. Upper panel: ACFs of the differentially resampled sgzK galaxies. The limiting magnitudes of each sgzK subsample are $22.0 < K \leq 23.0$, $21.0 < K \leq 22.0$, and $18.0 \leq K \leq 21.0$ (red, green and blue circles). Lower panel: The bias parameters of sgzKs.

4.2 Differential Luminosity Subsample

In this section, we show the results of clustering analysis on subsamples with different luminosities. Figure 11 shows the ACFs of sgzKs with different luminosities. The bin size of the brightest sgzK subsample ($18.0 \leq K \leq 21.0$) was increased to ($\delta \log(\theta) = 0.4$) to increase the S/N ratio. The error of each data point was larger than that for the cumulative subsamples due to the smaller sample size; however, these ACFs clearly show dependence on the K -band magnitudes and have the apparent excesses at small angular scale, as was the case for the cumulative subsamples.

In the same manner as with the cumulative resampling, we fitted our sgzK subsamples using a power law to determine the amplitudes of the ACFs and the correlation lengths. These data are listed in Table 3.

Bielby et al. (2014) showed that the correlation length of a star-forming galaxy depends on the stellar mass, whereas the correlation length of a passive galaxy does not (or is only weakly dependent of the stellar mass), in combination with the results of Coil et al. (2008), Bielby et al. (2010), and McCracken et al. (2010). Figure 12 shows the relationship between the correlation length and the stellar mass for both this study and previous studies (Bielby et al. 2014; Béthermin et al. 2014). We note that the stellar masses of our samples are estimated by their ($z - K$) colours and K -band magnitudes (see Section 4.3), whereas Béthermin et al. (2014) and Bielby et al. (2014) estimated by the SED fitting. The dependence of our correlation lengths on the stellar mass is consistent with previous studies.

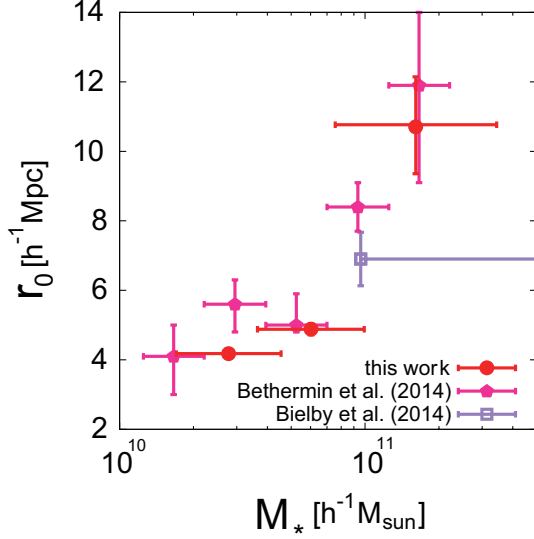


Figure 12. The dependence of the correlation length upon the stellar mass of *sgzKs/sBzKs*. Our measurements (red) show the good agreement with Béthermin et al. (2014) and Bielby et al. (2014) (magenta and purple, respectively). All correlation lengths are in units of $h^{-1}\text{Mpc}$, where $h = 0.7$.

4.3 Dark Halo Mass Estimation by the Large-Scale Clustering of *sgzK* Galaxies

We calculated the dark halo mass residing in our *sgzK* samples using the results of our accurate clustering analysis. We used the analytical model proposed by Sheth et al. (2001b) and Mo & White (2002), which has been shown to provide a connection between the bias of the dark halo and the mass of the dark halo based upon ellipsoidal collapse (Sheth et al. 2001a), provided bias of dark halo and the redshift are given.

We determined the amplitude of the two-point correlation function at $r = 8h^{-1}\text{Mpc}$, $\xi_{\text{sgzK}}(8h^{-1}\text{Mpc})$, assuming a power-law form of equation (22) and the correlation lengths. We define the effective bias parameter b_{eff} as

$$b_{\text{eff}} = \sqrt{\frac{\xi_{\text{sgzK}}(8h^{-1}\text{Mpc})}{\xi_{\text{DM}}(8h^{-1}\text{Mpc})}}, \quad (26)$$

where $\xi_{\text{DM}}(8h^{-1}\text{Mpc})$ is the model predicted correlation function of dark matter at an effective redshift with a scale of $8h^{-1}\text{Mpc}$ of each subsample (Hamana et al. 2002, and the references therein).

The bias of the dark halo can be calculated by the analytic formulae given by Sheth et al. (2001a) and Mo & White (2002). Here, we give a summary of these models. We define the linear growth factor $D(z)$ as

$$D(z) \equiv \frac{g(z)}{g(0)(1+z)}, \quad (27)$$

where $g(z)$ depends on the cosmological parameters, i.e.,

$$g(z) \approx \frac{5}{2}\Omega_{\text{m}}(z)[\Omega_{\text{m}}^{4/7}(z) - \Omega_{\Lambda}(z)] + (1 + \Omega_{\text{m}}(z)/2)(1 + \Omega_{\Lambda}(z)/70)^{-1}. \quad (28)$$

$\Omega_{\text{m}}(z)$ and $\Omega_{\Lambda}(z)$ are the cosmological parameters at a given

redshift, and they satisfy

$$\begin{aligned} \Omega_{\text{m}}(z) &= \frac{\Omega_{\text{m}}(1+z)^3}{E^2(z)} \\ \Omega_{\Lambda}(z) &= \frac{\Omega_{\Lambda}}{E^2(z)}, \end{aligned} \quad (29)$$

where $E(z)$ is given by

$$E(z) = [\Omega_{\Lambda} + (1 - \Omega_{\text{m}} - \Omega_{\Lambda})(1+z)^2 + \Omega_{\text{m}}(1+z)^3]^{1/2}. \quad (30)$$

The bias of the dark halo with a mass of M , $b(M)$, is given by

$$b(M) = 1 + \frac{1}{\delta_c}[\nu^2 + b\nu^{2(1-c)} - \frac{\nu^2/\sqrt{a}}{\nu^{2c} + b(1-c)(1-c/2)}], \quad (31)$$

where ν is defined by $\nu \equiv \sqrt{a}\delta_c/D(z)\sigma(M)$ and we assume the constants to be $a = 0.707$, $b = 0.5$, $c = 0.6$, and $\delta_c = 1.69$ (Sheth et al. 2001a). The term $\sigma(M)$ represents the root mean square of the density field fluctuation with a mass scale M , i.e.,

$$\sigma^2(M) = \frac{1}{2\pi^2} \int_0^\infty k^2 P(k) \tilde{W}(kR)^2 dk \quad (32)$$

and the redshift dependence is included via the linear growth factor, $D(z)$. The term $P(k)$ is the matter power spectrum and has a power-law index of $n = 1$. We calculated the power spectrum using the **CAMB** package (Lewis et al. 2000; Challinor & Lewis 2011). The term \tilde{W} is a window function, i.e.,

$$\tilde{W}(kR) = 3 \frac{\sin(kR) - kR\cos(kR)}{(kR)^3}, \quad (33)$$

and R is a Lagrangian radius of the halo, which gives the halo mass dependence, i.e.,

$$R(M) \equiv \left(\frac{3M}{4\pi\bar{\rho}_0}\right)^{1/3}, \quad (34)$$

where $\bar{\rho}_0$ represents the current mean density of the Universe.

We estimate the mean dark halo mass $\langle M_h \rangle$ and the minimum dark halo mass M_{min} from a comparison between the effective bias parameter and the bias of dark halo (see also Hayashi et al. 2007). We define the mean dark halo mass when the bias of dark halo with a mass of $\langle M_h \rangle$ is equal to the effective bias, i.e.,

$$b_{\text{eff}} = b(\langle M_h \rangle), \quad (35)$$

whereas the minimum dark halo mass is defined that the effective bias from equation (26) is equal to another expression of the effective bias as,

$$b_{\text{eff}} = \frac{\int_{M_{\text{min}}}^\infty b(M)n(M)dM}{\int_{M_{\text{min}}}^\infty n(M)dM}, \quad (36)$$

where $n(M)dM$ is the halo mass function.

Table 3 lists the dark halo masses. Our method to calculate the dark halo mass is based on the wide survey area and can thus be expected to be accurate, as the results depend strongly on the clustering signals at large angular scales. Our measurements satisfy $\langle M_h \rangle \approx 3M_{\text{min}}$ over almost all limiting magnitudes, which is consistent with the results of Hayashi et al. (2007), who reported that the mean dark halo mass is mainly determined by the less massive haloes, which

are more numerous than the more massive ones. It should be noted that this measurement method of using large-scale clustering assumes that each dark halo must contain a single galaxy. This assumption may be erroneous, however, as massive haloes have been reported to contain multiple galaxies (e.g., galaxy groups/clusters in the local Universe), whereas less massive dark haloes may contain no galaxy at all. A more precise halo model, such as HOD, which describes the number of galaxies in the dark halo as a function of the mass of halo, may be beneficial to provide a more detailed description of the structure model.

Bielby et al. (2014) and Béthermin et al. (2014) determined the dark halo mass by this method, though they divided their sample by the stellar mass. The $(z - K)$ colour and K -band magnitude allowed us to estimate the galaxy stellar mass by using the galaxy model of Koyama et al. (2013, and the references therein). The conversion equation from the $(z - K)$ and K -band magnitudes of sgzKs to stellar mass is given by

$$\begin{aligned} \log(M_*/10^{11}M_\odot) &= -0.4 \times (K - 21.90) \\ &+ (0.086 - 1.28 \times \exp(-0.921 \times (z - K))). \end{aligned} \quad (37)$$

We note that this was derived by assuming the Salpeter IMF (Salpeter 1955) and that the formation redshift is $z_f = 5$. The scatter in the stellar mass of each galaxy was ≈ 0.3 dex. We estimated the stellar mass of each sgzK using this equation and determined the average stellar mass, together with the standard deviation of each subsample.

We find that the minimum mass of the dark halo that resides in sgzKs satisfying $18.0 \leq K \leq 23.0$ is $M_{\min} = (4.60 \pm 0.38) \times 10^{11}h^{-1}M_\odot$ and the mean halo mass is $\langle M_h \rangle = (1.23 \pm 0.10) \times 10^{12}h^{-1}M_\odot$, which are approximately three times more massive than Hayashi et al. (2007) reported for $K < 23.2$. Additionally, Blanc et al. (2008) reported a minimum dark halo mass down to $K \lesssim 22.0$ of $M_{\min} \approx 3 \times 10^{12}h^{-1}M_\odot$, which is also more massive than our estimation, i.e., $M_{\min} = (8.45^{+1.75}_{-1.50}) \times 10^{11}h^{-1}M_\odot$. These inconsistencies may be caused by the shallowness of their K -band photometry data and the small sample size. We also compared our results with those of Bielby et al. (2014) and Béthermin et al. (2014) by calculating the stellar mass of each subsample. Our results were in good agreement in terms of the mean halo mass reported by Bielby et al. (2014) and Béthermin et al. (2014) over the entire stellar mass range. However, Bielby et al. (2014) only estimated the dark halo mass of the massive galaxies, $M_* \sim 10^{11}h^{-1}M_\odot$, and the error bars of the dark halo mass given by Béthermin et al. (2014) were relatively large. Here, however, the dark halo masses of $z \sim 2$ galaxies were determined for a wide range of K -band luminosities (and stellar masses), and we are able to provide relatively small error bars.

As discussed above, we succeeded in calculating accurate dark halo masses using large-scale clustering via our high-quality ACFs; however, it is essential to discuss the relationship between galaxy stellar masses and dark halo masses to reveal the star-formation history of galaxies. In our forthcoming paper, we will discuss the evolution of the dark halo mass, the number of satellite galaxies in the dark halo, and the relationship between the dark halo mass and the stellar mass using our accurate dark halo mass data

(including the HOD-derived dark halo mass and the galaxy stellar masses) to describe the evolution history of the galaxies.

5 CONCLUSIONS

We have described the colour-correction method, the number count, and the clustering properties of sgzK galaxies. Using the g -band data instead of B -band data, we succeeded in applying the gzK selection method to a largest-ever wide field. Our survey area was 5.2 deg^2 (4.8 deg^2 after masking), which is ~ 2.5 times larger than the widest BzK galaxy survey reported previously (McCracken et al. 2010).

The conclusions of this paper can be summarized as follows.

(i) Based on wide-field observation, we collected 41,112 sgzK galaxies, which is approximately twice the number of sBzK galaxies reported by McCracken et al. (2010). Because of the large number of galaxy samples, we obtained the high-quality ACFs and report a more accurate clustering analysis compared with previous BzK clustering studies. Our ACFs show an apparent excess from the power law, which is indicated by the halo model. We resampled our large number of sgzK sample into cumulative luminosity subsamples and differential luminosity subsamples to investigate the dependence of sgzK clustering properties on the K -band magnitude.

(ii) We found that the amplitudes of the ACFs determined using our cumulative luminosity subsamples were larger than those of previous studies. The amplitude of the ACFs is determined using the large-scale clustering of galaxies, and our ACFs have particularly large signals because of the wide survey area. This is the principal reason that the amplitudes of our ACFs may be considered more accurate than those of previous studies.

(iii) We calculated correlation lengths using the amplitudes of the ACFs, which are in good agreement with the results of previous studies, except for Hartley et al. (2008). The error bars of our correlation lengths are small compared with those of previous studies because of the high S/N ratio.

(iv) We found that correlation lengths calculated using our ACFs are dependent on the K -magnitude, i.e., brighter sgzKs in the K -band have a large correlation length, which is consistent with the results of many previous studies. It follows that bright sgzKs reside in massive dark haloes, which shows a agreement with the correlation between the correlation length and the dark halo mass inferred by Hayashi et al. (2007) and Béthermin et al. (2014).

(v) The dark halo masses were estimated from the large-scale sgzK clustering with small error bars and confirmed that our dark halo masses are in agreement with those of Bielby et al. (2014) and Béthermin et al. (2014).

ACKNOWLEDGMENTS

We are obligated to Masao Hayashi to his useful advice on this study. We also thank to Yusei Koyama to give us the conversion equation of stellar mass by using his model. This work is based on data collected at the Subaru Telescope, which is operated by the National Astronomical Observatory

of Japan (NAOJ). This research was supported by the Japan Society for the Promotion of Science through Grant-in-Aid for Scientific Research 23340050.

This research is based in part on data obtained as part of the United Kingdom Infra-Red Telescope (UKIRT) Deep Sky Survey. This study is based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada–France–Hawaii Telescope (CFHT), which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l’Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at TERAPIX and the Canadian Astronomy Data Center as part of the Canada–France–Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS.

REFERENCES

- Adelberger, K. L., Steidel, C. C., Pettini, M., et al. 2005, *ApJ*, 619, 697
- Arnouts, S., Walcher, C. J., Le Fèvre, O., et al. 2007, *A&A*, 476, 137
- Baba, H., Yasuda, N., Ichikawa, S.-I., et al. 2002, *Astronomical Data Analysis Software and Systems XI*, 281, 298
- Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393
- Béthermin, M., Kilbinger, M., Daddi, E., et al. 2014, *A&A*, 567, A103
- Bielby, R. M., Finoguenov, A., Tanaka, M., et al. 2010, *A&A*, 523, A66
- Bielby, R., Hudelot, P., McCracken, H. J., et al. 2012, *A&A*, 545, A23
- Bielby, R. M., Gonzalez-Perez, V., McCracken, H. J., et al. 2014, *A&A*, 568, A24
- Blanc, G. A., Lira, P., Barrientos, L. F., et al. 2008, *ApJ*, 681, 1099
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, 686, 1503
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Burgarella, D., Buat, V., Gruppioni, C., et al. 2013, *A&A*, 554, A70
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, 533, 682
- Carpenter, J. M. 2001, *AJ*, 121, 2851
- Carroll, S. M., Press, W. H., & Turner, E. L. 1992, *ARAA*, 30, 499
- Coil, A. L., Newman, J. A., Croton, D., et al. 2008, *ApJ*, 672, 153
- Cowie, L. L., Songaila, A., & Barger, A. J. 1999, *AJ*, 118, 603
- Challinor, A., & Lewis, A. 2011, *Astrophysics Source Code Library*, 5013
- Daddi, E., Cimatti, A., Renzini, A., et al. 2004, *ApJ*, 617, 746
- Dickinson, M., Papovich, C., Ferguson, H. C., & Budavári, T. 2003, *ApJ*, 587, 25
- Efstathiou, G., Bernstein, G., Tyson, J. A., Katz, N., & Guhathakurta, P. 1991, *ApJL*, 380, L47
- Fang, G., Kong, X., Chen, Y., & Lin, X. 2012, *ApJ*, 751, 109
- Ferguson, H. C., Dickinson, M., Giavalisco, M., et al. 2004, *ApJL*, 600, L107
- Gwyn, S. D. J. 2011, arXiv:1101.1084
- Hamana, T., Yoshida, N., & Suto, Y. 2002, *ApJ*, 568, 455
- Hamana, T., Ouchi, M., Shimasaku, K., Kayo, I., & Suto, Y. 2004, *MNRAS*, 347, 813
- Hartley, W. G., Lane, K. P., Almaini, O., et al. 2008, *MNRAS*, 391, 1301
- Hayashi, M., Shimasaku, K., Motohara, K., et al. 2007, *ApJ*, 660, 72
- Hildebrandt, H., Pielorz, J., Erben, T., et al. 2009, *A&A*, 498, 725
- Hopkins, A. M., & Beacom, J. F. 2006, *ApJ*, 651, 142
- Ilbert, O., Salvato, M., Le Floc’h, E., et al. 2010, *ApJ*, 709, 644
- Jose, C., Srianand, R., & Subramanian, K. 2013, *MNRAS*, 435, 368
- Kashikawa, N., Yoshida, M., Shimasaku, K., et al. 2006, *ApJ*, 637, 631
- Kashikawa, N., Ishizaki, Y., Willott, C. J., et al. 2015, *ApJ*, 798, 28
- Kodama, T., Arimoto, N., Barger, A. J., & Arag’ón-Salamanca, A. 1998, *A&A*, 334, 99
- Kodama, T., Bell, E. F., & Bower, R. G. 1999, *MNRAS*, 302, 152
- Kong, X., Daddi, E., Arimoto, N., et al. 2006, *ApJ*, 638, 72
- Koyama, Y., Smail, I., Kurk, J., et al. 2013, *MNRAS*, 434, 423
- Landy, S. D., & Szalay, A. S. 1993, *ApJ*, 412, 64
- Lane, K. P., Almaini, O., Foucaud, S., et al. 2007, *MNRAS*, 379, L25
- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *MNRAS*, 379, 1599
- Lewis, A., Challinor, A., & Lasenby, A. 2000, *ApJ*, 538, 473
- Limber, D. N. 1953, *ApJ*, 117, 134
- McCracken, H. J., Capak, P., Salvato, M., et al. 2010, *ApJ*, 708, 202
- Mo, H. J., & White, S. D. M. 1996, *MNRAS*, 282, 347
- Mo, H. J., & White, S. D. M. 2002, *MNRAS*, 336, 112
- Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, *ApJS*, 206, 8
- Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, *ApJL*, 660, L43
- Ouchi, M., Shimasaku, K., Okamura, S., et al. 2004a, *ApJ*, 611, 660
- Ouchi, M., Shimasaku, K., Okamura, S., et al. 2004b, *ApJ*, 611, 685
- Palanque-Delabrouille, N., Magneville, C., Yèche, C., et al. 2013, *A&A*, 551, A29
- Peebles, P. J. E. 1980, *Research supported by the National Science Foundation*. Princeton, N.J., Princeton University Press, 1980. 435 p.,
- Pérez-González, P. G., Rieke, G. H., Villar, V., et al. 2008, *ApJ*, 675, 234
- Roche, N., & Eales, S. A. 1999, *MNRAS*, 307, 703
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Santini, P., Fontana, A., Grazian, A., et al. 2009, *A&A*, 504, 751
- Sato, T., Sawicki, M., & Arcila-Osejo, L. 2014, *MNRAS*, 443, 2661
- Sheth, R. K., & Tormen, G. 1999, *MNRAS*, 308, 119
- Sheth, R. K., Mo, H. J., & Tormen, G. 2001, *MNRAS*, 323,

1

- Sheth, R. K., Diaferio, A., Hui, L., & Scoccimarro, R. 2001, MNRAS, 326, 463
- Shimasaku, K., Ouchi, M., Furusawa, H., et al. 2005, PASJ, 57, 447
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
- Yagi, M., Kashikawa, N., Sekiguchi, M., et al. 2002, AJ, 123, 66
- Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, ApJ, 630, 1
- Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59
- Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, ApJ, 633, 791
- Zheng, Z., Coil, A. L., & Zehavi, I. 2007, ApJ, 667, 760

Table 3. The detailed data of cumulative/differential subsamples

	N(sgzK)	z_{eff}	$A_{\omega}(1^{\circ}) \times 10^{-3}$	$r_0 [h^{-1} \text{ Mpc}]$	b_{eff}	$M_{\text{min}} [h^{-1} \text{ M}_{\odot}]$	$\langle M_h \rangle [h^{-1} \text{ M}_{\odot}]$	$M_{\star} [h^{-1} \text{ M}_{\odot}]$
$18.0 \leq K \leq 21.0$	2,086	1.43	18.48 ± 4.57	10.52 ± 1.30	4.16 ± 0.46	$(1.71^{+0.76}_{-0.60}) \times 10^{13}$	$(3.26^{+1.23}_{-1.02}) \times 10^{13}$	$(1.61^{+1.83}_{-0.85}) \times 10^{11}$
$18.0 \leq K \leq 21.5$	6,243	1.61	7.23 ± 1.40	5.61 ± 0.57	2.36 ± 0.21	$(1.84^{+0.97}_{-0.63}) \times 10^{12}$	$(4.39^{+1.82}_{-1.45}) \times 10^{12}$	$(1.24^{+1.00}_{-0.55}) \times 10^{11}$
$18.0 \leq K \leq 22.0$	15,247	1.68	4.36 ± 0.27	4.67 ± 0.17	2.00 ± 0.07	$(8.45^{+1.75}_{-1.50}) \times 10^{11}$	$(2.14^{+0.27}_{-0.31}) \times 10^{12}$	$(8.64^{+7.00}_{-3.87}) \times 10^{10}$
$18.0 \leq K \leq 22.5$	29,158	1.73	3.83 ± 0.14	4.47 ± 0.08	1.92 ± 0.03	$(6.78^{+0.54}_{-0.50}) \times 10^{11}$	$(1.78^{+0.14}_{-0.13}) \times 10^{12}$	$(6.13^{+5.56}_{-2.92}) \times 10^{10}$
$18.0 \leq K \leq 23.0$	41,112	1.76	3.33 ± 0.09	4.12 ± 0.07	1.79 ± 0.03	$(4.60^{+0.38}_{-0.38}) \times 10^{11}$	$(1.23^{+0.10}_{-0.10}) \times 10^{12}$	$(4.95^{+4.80}_{-2.44}) \times 10^{10}$
$18.0 \leq K \leq 21.0$	2,086	1.43	18.48 ± 4.57	10.52 ± 1.30	4.16 ± 0.46	$(1.71^{+0.76}_{-0.60}) \times 10^{13}$	$(3.26^{+1.23}_{-1.02}) \times 10^{13}$	$(1.61^{+1.83}_{-0.85}) \times 10^{11}$
$21.0 < K \leq 22.0$	13,163	1.65	4.81 ± 0.25	4.88 ± 0.26	2.08 ± 0.06	$(1.07^{+0.15}_{-0.17}) \times 10^{12}$	$(2.55^{+0.34}_{-0.32}) \times 10^{12}$	$(6.02^{+3.91}_{-2.37}) \times 10^{10}$
$22.0 < K \leq 23.0$	25,865	1.73	3.32 ± 0.09	4.18 ± 0.07	1.81 ± 0.03	$(4.93^{+0.44}_{-0.43}) \times 10^{11}$	$(1.32^{+0.09}_{-0.12}) \times 10^{12}$	$(2.80^{+1.77}_{-1.08}) \times 10^{10}$